

Statistics for Economists II

Lecture 11: Testing Means & Variances — Two-Population Problems

Güneş A. Aşık

Department of Economics
Middle East Technical University

Spring 2025–2026

Roadmap for Today

- 1 Motivation: Comparing Populations
- 2 Testing the Variance of a Normal Population
- 3 Two-Population Tests for Means
 - Paired t -Test
 - Independent Samples: Pooled t -Test
 - Welch's t -Test
- 4 Two-Population Tests for Variances: The F -Test
- 5 Summary of Tests
- 6 Practice Problems
- 7 Key Takeaways & Further Reading

Why Compare Two Populations?

- Most economic questions are **comparative**:
 - Do men and women earn the **same**?
 - Did the policy make growth **higher**?
 - Is market A **more volatile** than market B?

Comparing two groups is the **bread and butter** of empirical economics. Every randomised controlled trial, every diff-in-diff, every wage-gap study starts with a two-sample comparison.

Why Compare Two Populations?

- Most economic questions are **comparative**:
 - Do men and women earn the **same**?
 - Did the policy make growth **higher**?
 - Is market A **more volatile** than market B?
- Single-population tests (Weeks 9–10) answer: “Is $\mu = \mu_0$?”

Comparing two groups is the **bread and butter** of empirical economics. Every randomised controlled trial, every diff-in-diff, every wage-gap study starts with a two-sample comparison.

Why Compare Two Populations?

- Most economic questions are **comparative**:
 - Do men and women earn the **same**?
 - Did the policy make growth **higher**?
 - Is market A **more volatile** than market B?
- Single-population tests (Weeks 9–10) answer: “Is $\mu = \mu_0$?”
- Today: “Is $\mu_1 = \mu_2$?” and “Is $\sigma_1^2 = \sigma_2^2$?”

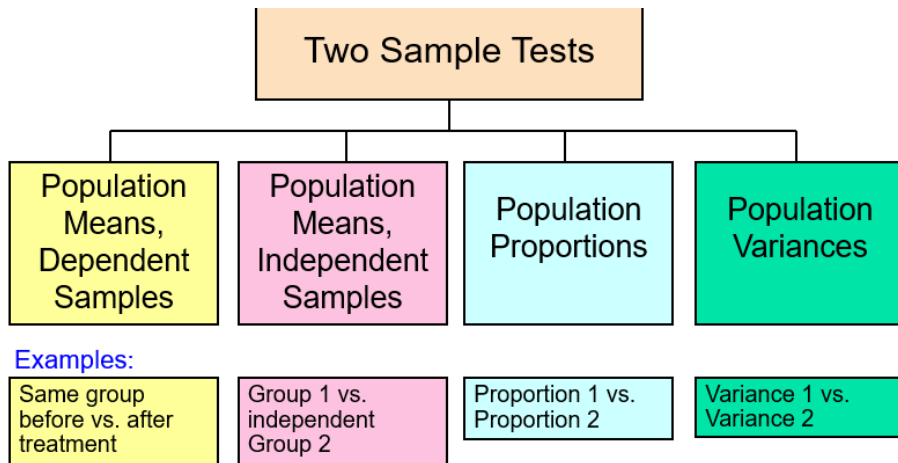
Comparing two groups is the **bread and butter** of empirical economics. Every randomised controlled trial, every diff-in-diff, every wage-gap study starts with a two-sample comparison.

Why Compare Two Populations?

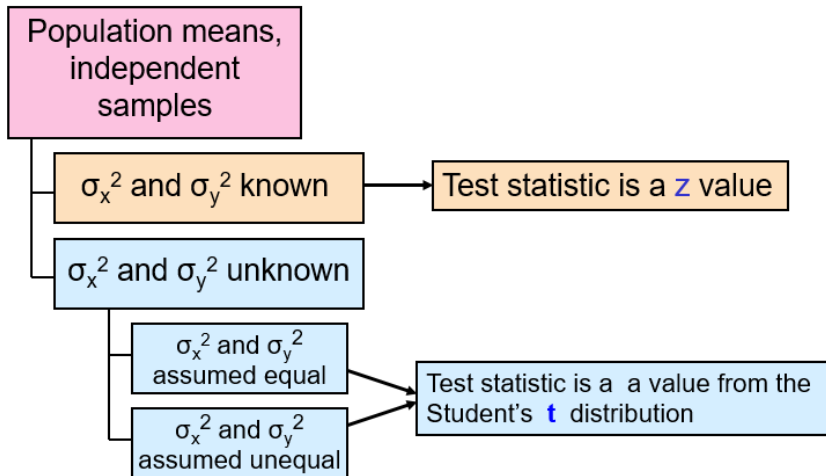
- Most economic questions are **comparative**:
 - Do men and women earn the **same**?
 - Did the policy make growth **higher**?
 - Is market A **more volatile** than market B?
- Single-population tests (Weeks 9–10) answer: “Is $\mu = \mu_0$?”
- Today: “Is $\mu_1 = \mu_2$?” and “Is $\sigma_1^2 = \sigma_2^2$?”
- We also complete the one-population toolkit with the **chi-squared test for variance**

Comparing two groups is the **bread and butter** of empirical economics. Every randomised controlled trial, every diff-in-diff, every wage-gap study starts with a two-sample comparison.

Today's Big Picture



Today's Big Picture



The Chi-Squared Test for Variance

Setup

Let $X_1, \dots, X_n \stackrel{\text{iid}}{\sim} N(\mu, \sigma^2)$, both unknown. We wish to test:

$$H_0: \sigma^2 = \sigma_0^2 \quad \text{vs.} \quad H_1: \sigma^2 \neq \sigma_0^2$$

Test statistic:

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2} \sim \chi^2(n-1) \quad \text{under } H_0$$

Under H_0 , the sample variance S^2 should be “close” to σ_0^2 , so the ratio $(n-1)S^2/\sigma_0^2$ should behave like a $\chi^2(n-1)$ random variable. Values far in either tail signal that $\sigma^2 \neq \sigma_0^2$.

Chi-Squared Test: Rejection Regions

Two-sided: $H_1: \sigma^2 \neq \sigma_0^2$

Reject H_0 if $\chi^2 < \chi_{1-\alpha/2, n-1}^2$ or $\chi^2 > \chi_{\alpha/2, n-1}^2$

One-sided (upper): $H_1: \sigma^2 > \sigma_0^2$

Reject H_0 if $\chi^2 > \chi_{\alpha, n-1}^2$

One-sided (lower): $H_1: \sigma^2 < \sigma_0^2$

Reject H_0 if $\chi^2 < \chi_{1-\alpha, n-1}^2$

Important

- The χ^2 distribution is **right-skewed** and takes only non-negative values
- The two-sided critical values are **not symmetric**
- This test **requires normality**—it is sensitive to non-normal populations

Example: Quality Control in Manufacturing

Problem: A factory claims that the variance of widget diameter is at most $\sigma_0^2 = 0.04 \text{ cm}^2$. An inspector samples $n = 20$ widgets and finds $s^2 = 0.065$. Test at $\alpha = 0.05$.

Example: Quality Control in Manufacturing

Problem: A factory claims that the variance of widget diameter is at most $\sigma_0^2 = 0.04 \text{ cm}^2$. An inspector samples $n = 20$ widgets and finds $s^2 = 0.065$. Test at $\alpha = 0.05$.

Solution:

- $H_0: \sigma^2 = 0.04$ vs. $H_1: \sigma^2 > 0.04$ (one-sided upper)

- Test statistic:

$$\chi^2 = \frac{(20 - 1)(0.065)}{0.04} = \frac{19 \times 0.065}{0.04} = 30.875$$

- Critical value: $\chi_{0.05, 19}^2 = 30.14$

- Since $30.875 > 30.14$, we **reject** H_0

The observed sample variance is significantly larger than the claimed standard. The inspector has statistical evidence that the production line is **too variable**—quality control intervention is warranted.

Two-Population Mean Tests: The Central Question

General Framework

We have two populations with means μ_1 and μ_2 . The most common test:

$$H_0: \mu_1 - \mu_2 = \delta_0 \quad (\text{often } \delta_0 = 0)$$

The key question: **Are the samples paired or independent?**

Paired (Dependent)

- Same subjects measured twice
- Matched subjects (e.g., twins)
- Before/after comparisons

Independent

- Two separate random samples
- Different subjects in each group
- E.g., men vs. women

The Paired t -Test

Setup

Observe (X_{1i}, X_{2i}) for $i = 1, \dots, n$. Define **differences**: $D_i = X_{1i} - X_{2i}$.
If $D_i \sim N(\mu_D, \sigma_D^2)$, then testing $H_0: \mu_1 = \mu_2$ reduces to testing $H_0: \mu_D = 0$.

Test statistic:

$$t = \frac{\bar{D} - 0}{S_D / \sqrt{n}} \sim t(n-1) \quad \text{under } H_0$$

where $\bar{D} = \frac{1}{n} \sum_{i=1}^n D_i$ and $S_D^2 = \frac{1}{n-1} \sum_{i=1}^n (D_i - \bar{D})^2$.

By computing **differences**, we reduce a two-sample problem to a **one-sample** t -test! The pairing “controls for” individual-level heterogeneity, making the test more powerful when subjects vary a lot.

Example: Policy Evaluation—Job Training Program

Problem: A government evaluates a job-training program. Monthly earnings (\$1000s) of $n = 8$ workers are recorded **before** and **after** the program:

Worker	1	2	3	4	5	6	7	8
Before	2.1	3.4	2.8	1.9	3.0	2.5	2.2	3.1
After	2.8	3.5	3.2	2.5	3.4	2.9	2.6	3.6
D_i (After–Before)	0.7	0.1	0.4	0.6	0.4	0.4	0.4	0.5

Example: Policy Evaluation—Job Training Program

Problem: A government evaluates a job-training program. Monthly earnings (\$1000s) of $n = 8$ workers are recorded **before** and **after** the program:

Worker	1	2	3	4	5	6	7	8
Before	2.1	3.4	2.8	1.9	3.0	2.5	2.2	3.1
After	2.8	3.5	3.2	2.5	3.4	2.9	2.6	3.6
D_i (After–Before)	0.7	0.1	0.4	0.6	0.4	0.4	0.4	0.5

$$\bar{D} = 0.4375, \quad S_D = 0.1685$$

$$t = \frac{0.4375}{0.1685/\sqrt{8}} = \frac{0.4375}{0.0596} = 7.34$$

$t_{0.025, 7} = 2.365$. Since $7.34 > 2.365$, we **reject** H_0 at $\alpha = 0.05$.

Conclusion: Strong evidence that the training program **increased** earnings.

When to Use Paired vs. Independent Tests

Use Paired When...

- Same individual measured at two time points (before/after)
- Matched subjects (twins, siblings, spouses)
- Each observation in group 1 has a natural “partner” in group 2
- Two treatments applied to the same experimental unit

Use Independent When...

- Two separate, unrelated groups
- No natural pairing exists
- Sample sizes may differ ($n_1 \neq n_2$)
- E.g., wages of men vs. women from different surveys

Common Mistake

Using an **independent** test when data are paired wastes information and **reduces power**. If the pairing exists, always use the paired test!

Independent Samples: Equal Variances (Pooled t -Test)

Setup

$X_{11}, \dots, X_{1n_1} \stackrel{\text{iid}}{\sim} N(\mu_1, \sigma^2)$ and $X_{21}, \dots, X_{2n_2} \stackrel{\text{iid}}{\sim} N(\mu_2, \sigma^2)$, **independent** samples.

$H_0: \mu_1 - \mu_2 = \delta_0$ (typically $\delta_0 = 0$)

Pooled variance estimator:

$$S_p^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$$

Test statistic:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim t(n_1 + n_2 - 2) \quad \text{under } H_0$$

The pooled estimator **combines** information from both samples to get a better estimate of the common variance σ^2 . It is a weighted average of S_1^2 and S_2^2 with weights proportional to degrees of freedom.

Independent Samples: Unequal Variances (Welch's t -Test)

Setup

Same as before, but now $\sigma_1^2 \neq \sigma_2^2$ (or we do not want to assume equality).

Test statistic:

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \sim t(\nu) \quad (\text{approximately})$$

Welch–Satterthwaite degrees of freedom:

$$\nu = \frac{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2} \right)^2}{\frac{(S_1^2/n_1)^2}{n_1 - 1} + \frac{(S_2^2/n_2)^2}{n_2 - 1}}$$

(Round ν down to the nearest integer.)

Example: Gender Wage Gap

Problem: An economist collects hourly wages (in TL) from two independent samples:

- Men: $n_1 = 40$, $\bar{x}_1 = 58.5$, $s_1 = 12.3$
- Women: $n_2 = 35$, $\bar{x}_2 = 52.8$, $s_2 = 10.1$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ using Welch's test.

Example: Gender Wage Gap

Problem: An economist collects hourly wages (in TL) from two independent samples:

- Men: $n_1 = 40$, $\bar{x}_1 = 58.5$, $s_1 = 12.3$
- Women: $n_2 = 35$, $\bar{x}_2 = 52.8$, $s_2 = 10.1$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ using Welch's test.

Solution:

$$t = \frac{58.5 - 52.8}{\sqrt{\frac{12.3^2}{40} + \frac{10.1^2}{35}}} = \frac{5.7}{\sqrt{3.782 + 2.915}} = \frac{5.7}{2.587} = 2.203$$

Example: Gender Wage Gap

Problem: An economist collects hourly wages (in TL) from two independent samples:

- Men: $n_1 = 40$, $\bar{x}_1 = 58.5$, $s_1 = 12.3$
- Women: $n_2 = 35$, $\bar{x}_2 = 52.8$, $s_2 = 10.1$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ using Welch's test.

Solution:

$$t = \frac{58.5 - 52.8}{\sqrt{\frac{12.3^2}{40} + \frac{10.1^2}{35}}} = \frac{5.7}{\sqrt{3.782 + 2.915}} = \frac{5.7}{2.587} = 2.203$$

$$\text{Welch df: } \nu = \frac{(3.782 + 2.915)^2}{\frac{3.782^2}{39} + \frac{2.915^2}{34}} = \frac{44.85}{0.367 + 0.250} = 72.7 \approx 72$$

$t_{0.025, 72} \approx 1.993$. Since $2.203 > 1.993$, we **reject** H_0 .

Conclusion: Statistically significant evidence of a **gender wage gap** at the 5% level.

Large-Sample Approximation (z-Test)

When Both Samples Are Large

If n_1 and n_2 are both large (say ≥ 30), the CLT gives:

$$z = \frac{(\bar{X}_1 - \bar{X}_2) - \delta_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \underset{\text{approx}}{\sim} N(0, 1)$$

No need for the t -distribution or the Welch–Satterthwaite correction.

For large samples, the t -distribution converges to the standard normal, so the z -test and Welch's t -test give essentially the **same answer**. But for small samples ($n < 30$), the t -distribution is more appropriate because it has heavier tails and is more conservative.

Large-Sample Approximation (z-Test)

Normality Assumption

With large samples, the CLT justifies the normal approximation even if the underlying populations are **not** normal. With small samples, you need normality (or at least approximate symmetry).

Pooled vs. Welch vs. z: Quick Comparison

Feature	Pooled t	Welch t	z-test
Assumes $\sigma_1^2 = \sigma_2^2$?	Yes	No	No
Requires normality?	Yes	Yes	No (CLT)
Sample size requirement	Small OK	Small OK	$n_1, n_2 \geq 30$
Degrees of freedom	$n_1 + n_2 - 2$	Welch–Satt. ν	∞ (use z)
When to use?	Equal var. known	Default choice	Large samples

In practice, most economists use **Welch's test** as the default because it is robust to unequal variances and performs well even when variances *are* equal. If sample sizes are large, all three give nearly identical results.

The F -Test for Equality of Variances

Setup

Two independent normal samples: $X_{1i} \sim N(\mu_1, \sigma_1^2)$, $i = 1, \dots, n_1$ and $X_{2j} \sim N(\mu_2, \sigma_2^2)$, $j = 1, \dots, n_2$.

$$H_0: \sigma_1^2 = \sigma_2^2 \quad \text{vs.} \quad H_1: \sigma_1^2 \neq \sigma_2^2$$

Test statistic:

$$F = \frac{S_1^2}{S_2^2} \sim F(n_1 - 1, n_2 - 1) \quad \text{under } H_0$$

If $\sigma_1^2 = \sigma_2^2$, then S_1^2 and S_2^2 estimate the same quantity, and their ratio should be close to **1**. A ratio far from 1 (either very large or very small) provides evidence against H_0 .

F-Test: Rejection Regions

Two-sided ($H_1: \sigma_1^2 \neq \sigma_2^2$):

Reject H_0 if $F > F_{\alpha/2, n_1-1, n_2-1}$ or $F < F_{1-\alpha/2, n_1-1, n_2-1}$

Convention: Place the **larger** sample variance in the numerator so $F \geq 1$, then use only the upper critical value $F_{\alpha/2}$ (and double the one-tail p -value).

One-sided ($H_1: \sigma_1^2 > \sigma_2^2$):

Reject H_0 if $F > F_{\alpha, n_1-1, n_2-1}$

Sensitivity to Non-Normality

The F -test for variances is **very sensitive** to departures from normality—much more so than t -tests for means.

Example: Comparing Investment Risk

Problem: An investor wants to compare the volatility of two portfolios. Monthly returns (%) are:

- Portfolio A: $n_1 = 25$, $s_1^2 = 18.6$
- Portfolio B: $n_2 = 21$, $s_2^2 = 9.2$

Test whether the variances are equal at $\alpha = 0.10$.

Example: Comparing Investment Risk

Problem: An investor wants to compare the volatility of two portfolios. Monthly returns (%) are:

- Portfolio A: $n_1 = 25$, $s_1^2 = 18.6$
- Portfolio B: $n_2 = 21$, $s_2^2 = 9.2$

Test whether the variances are equal at $\alpha = 0.10$.

Solution:

$$F = \frac{S_1^2}{S_2^2} = \frac{18.6}{9.2} = 2.022 \quad (S_1^2 > S_2^2 \text{ placed in numerator})$$

$F_{0.05, 24, 20} \approx 2.08$ (two-sided, so $\alpha/2 = 0.05$).

Since $2.022 < 2.08$, we **fail to reject** H_0 at $\alpha = 0.10$.

Even though Portfolio A has double the sample variance, with these sample sizes we do **not** have sufficient evidence to conclude the population variances differ. The F -test requires quite extreme ratios (or large samples) to reject—it has relatively low power.

F-Test as a Preliminary Step

The F -test is often used **before** a two-sample t -test to decide between pooled and Welch:

- 1 Test $H_0: \sigma_1^2 = \sigma_2^2$ using the F -test
- 2 If not rejected $\Rightarrow$ use **pooled** t -test
- 3 If rejected $\Rightarrow$ use **Welch's** t -test

Caveat

This two-step procedure inflates the overall Type I error. Many modern texts recommend **always using Welch's test** rather than using the F -test as a preliminary.

Think of the F -test as a diagnostic, not a decision rule. In econometrics, you will encounter more robust alternatives (e.g., heteroskedasticity-robust standard errors) that bypass the need to test for equal variances altogether.

Master Summary Table

Scenario	Test Statistic	Distribution	Key Assumption
One-pop. variance	$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$	$\chi^2(n-1)$	Normal pop.
Paired means	$t = \frac{\bar{D}}{S_D/\sqrt{n}}$	$t(n-1)$	Paired obs., normal D_i
Indep. means, $\sigma_1^2 = \sigma_2^2$	$t = \frac{\bar{X}_1 - \bar{X}_2}{S_p \sqrt{1/n_1 + 1/n_2}}$	$t(n_1 + n_2 - 2)$	Equal variances, normal
Indep. means, $\sigma_1^2 \neq \sigma_2^2$	$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	$t(\nu_{\text{Welch}})$	Normal pops.
Indep. means, large n	$z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}}$	$N(0, 1)$	Large samples (CLT)
Two-pop. variances	$F = \frac{S_1^2}{S_2^2}$	$F(n_1 - 1, n_2 - 1)$	Normal pops.

Practice Problem 1: Chi-Squared Test for Variance

Problem: A cereal company advertises that the standard deviation of the fill weight is at most 2 grams ($\sigma_0 = 2$, so $\sigma_0^2 = 4$). A consumer agency samples $n = 25$ boxes and finds $s^2 = 5.8$. Test $H_0: \sigma^2 \leq 4$ vs. $H_1: \sigma^2 > 4$ at $\alpha = 0.01$.

Practice Problem 1: Chi-Squared Test for Variance

Problem: A cereal company advertises that the standard deviation of the fill weight is at most 2 grams ($\sigma_0 = 2$, so $\sigma_0^2 = 4$). A consumer agency samples $n = 25$ boxes and finds $s^2 = 5.8$. Test $H_0: \sigma^2 \leq 4$ vs. $H_1: \sigma^2 > 4$ at $\alpha = 0.01$.

Solution:

- Test statistic: $\chi^2 = \frac{(25-1)(5.8)}{4} = \frac{139.2}{4} = 34.8$
- Critical value: $\chi_{0.01, 24}^2 = 42.98$
- Since $34.8 < 42.98$, we **fail to reject** H_0 at $\alpha = 0.01$

At the 1% significance level, there is **insufficient evidence** to conclude that the fill-weight variability exceeds the company's claim. Note that we would **reject at** $\alpha = 0.05$ since $\chi_{0.05, 24}^2 = 36.42 > 34.8$ —wait, actually $34.8 < 36.42$, so we would fail to reject at 5% as well. The sample variance is elevated but not extreme enough relative to the degrees of freedom.

Practice Problem 2: Paired t -Test

Problem: A study evaluates a new teaching method. Test scores of $n = 10$ students are recorded before and after the intervention:

Student	1	2	3	4	5	6	7	8	9	10
Before	72	68	85	60	78	55	90	63	74	81
After	78	74	88	67	82	62	93	70	80	85
D_i	6	6	3	7	4	7	3	7	6	4

Test $H_0: \mu_D = 0$ vs. $H_1: \mu_D > 0$ at $\alpha = 0.05$.

Practice Problem 2: Paired t -Test

Problem: A study evaluates a new teaching method. Test scores of $n = 10$ students are recorded before and after the intervention:

Student	1	2	3	4	5	6	7	8	9	10
Before	72	68	85	60	78	55	90	63	74	81
After	78	74	88	67	82	62	93	70	80	85
D_i	6	6	3	7	4	7	3	7	6	4

Test $H_0: \mu_D = 0$ vs. $H_1: \mu_D > 0$ at $\alpha = 0.05$.

Solution: $\bar{D} = 5.3$, $S_D = 1.636$

$$t = \frac{5.3}{1.636/\sqrt{10}} = \frac{5.3}{0.5174} = 10.24$$

$t_{0.05,9} = 1.833$. Since $10.24 \gg 1.833$, we **strongly reject** H_0 .

The new teaching method significantly **improves** test scores.

Practice Problem 3: Welch's t -Test

Problem: An economist compares per-capita spending in two cities:

- City A: $n_1 = 15$, $\bar{x}_1 = 2,450$, $s_1 = 380$
- City B: $n_2 = 18$, $\bar{x}_2 = 2,180$, $s_2 = 520$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ (do not assume equal variances).

Practice Problem 3: Welch's t -Test

Problem: An economist compares per-capita spending in two cities:

- City A: $n_1 = 15$, $\bar{x}_1 = 2,450$, $s_1 = 380$
- City B: $n_2 = 18$, $\bar{x}_2 = 2,180$, $s_2 = 520$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ (do not assume equal variances).

Solution:

$$t = \frac{2450 - 2180}{\sqrt{\frac{380^2}{15} + \frac{520^2}{18}}} = \frac{270}{\sqrt{9627 + 15,022}} = \frac{270}{\sqrt{24,649}} = \frac{270}{157.0} = 1.720$$

Practice Problem 3: Welch's t -Test

Problem: An economist compares per-capita spending in two cities:

- City A: $n_1 = 15$, $\bar{x}_1 = 2,450$, $s_1 = 380$
- City B: $n_2 = 18$, $\bar{x}_2 = 2,180$, $s_2 = 520$

Test $H_0: \mu_1 = \mu_2$ vs. $H_1: \mu_1 \neq \mu_2$ at $\alpha = 0.05$ (do not assume equal variances).

Solution:

$$t = \frac{2450 - 2180}{\sqrt{\frac{380^2}{15} + \frac{520^2}{18}}} = \frac{270}{\sqrt{9627 + 15,022}} = \frac{270}{\sqrt{24,649}} = \frac{270}{157.0} = 1.720$$

$$\text{Welch df: } \nu = \frac{24,649^2}{\frac{9627^2}{14} + \frac{15,022^2}{17}} = \frac{607.4 \times 10^6}{6.62 \times 10^6 + 13.28 \times 10^6} = 30.5 \approx 30$$

$t_{0.025, 30} = 2.042$. Since $1.720 < 2.042$, we **fail to reject** H_0 .

Conclusion: No significant difference in per-capita spending at $\alpha = 0.05$.

Practice Problem 4: F -Test for Variances

Problem: Two stock market indices are compared for volatility. Daily returns are:

- BIST 100: $n_1 = 31$, $s_1^2 = 4.82$
- DAX: $n_2 = 26$, $s_2^2 = 2.15$

Test $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 > \sigma_2^2$ at $\alpha = 0.05$.

Practice Problem 4: F -Test for Variances

Problem: Two stock market indices are compared for volatility. Daily returns are:

- BIST 100: $n_1 = 31$, $s_1^2 = 4.82$
- DAX: $n_2 = 26$, $s_2^2 = 2.15$

Test $H_0: \sigma_1^2 = \sigma_2^2$ vs. $H_1: \sigma_1^2 > \sigma_2^2$ at $\alpha = 0.05$.

Solution:

$$F = \frac{S_1^2}{S_2^2} = \frac{4.82}{2.15} = 2.242$$

$F_{0.05, 30, 25} \approx 1.92$ (one-sided upper).

Since $2.242 > 1.92$, we **reject** H_0 .

The BIST 100 exhibits **significantly greater volatility** than the DAX. This is consistent with the stylised fact that emerging-market indices tend to be more volatile than developed-market ones—a key consideration for international portfolio diversification.

Practice Problem 5: Choosing the Right Test

Problem: For each scenario, identify the appropriate test.

- (a) An analyst tests whether daily calorie intake ($\sigma = 400$) of adults in City X differs from 2,200. Sample: $n = 50$.

Practice Problem 5: Choosing the Right Test

Problem: For each scenario, identify the appropriate test.

- (a) An analyst tests whether daily calorie intake ($\sigma = 400$) of adults in City X differs from 2,200. Sample: $n = 50$.

Answer: One-sample z-test (known σ , n large).

- (b) Blood pressure of 15 patients is measured before and after taking a new drug.

Practice Problem 5: Choosing the Right Test

Problem: For each scenario, identify the appropriate test.

- (a) An analyst tests whether daily calorie intake ($\sigma = 400$) of adults in City X differs from 2,200. Sample: $n = 50$.

Answer: One-sample z-test (known σ , n large).

- (b) Blood pressure of 15 patients is measured before and after taking a new drug.

Answer: Paired t -test (same subjects, before/after).

- (c) Average spending of 40 online shoppers vs. 45 in-store shoppers is compared.

Practice Problem 5: Choosing the Right Test

Problem: For each scenario, identify the appropriate test.

- (a) An analyst tests whether daily calorie intake ($\sigma = 400$) of adults in City X differs from 2,200. Sample: $n = 50$.

Answer: One-sample z-test (known σ , n large).

- (b) Blood pressure of 15 patients is measured before and after taking a new drug.

Answer: Paired t -test (same subjects, before/after).

- (c) Average spending of 40 online shoppers vs. 45 in-store shoppers is compared.

Answer: Welch's t -test (independent samples, no reason to assume equal variances; z-test also acceptable given large n).

- (d) An investor tests whether the variance of a tech fund's returns equals that of a bond fund.

Practice Problem 5: Choosing the Right Test

Problem: For each scenario, identify the appropriate test.

- (a) An analyst tests whether daily calorie intake ($\sigma = 400$) of adults in City X differs from 2,200. Sample: $n = 50$.

Answer: One-sample z-test (known σ , n large).

- (b) Blood pressure of 15 patients is measured before and after taking a new drug.

Answer: Paired t -test (same subjects, before/after).

- (c) Average spending of 40 online shoppers vs. 45 in-store shoppers is compared.

Answer: Welch's t -test (independent samples, no reason to assume equal variances; z-test also acceptable given large n).

- (d) An investor tests whether the variance of a tech fund's returns equals that of a bond fund.

Answer: F -test for equality of two variances.

Key Takeaways

- 1 The **chi-squared test** lets us test hypotheses about a single population's variance—crucial for quality control and risk assessment
- 2 The **paired t -test** is powerful for before/after and matched-sample designs; it reduces a two-sample problem to a one-sample problem
- 3 For **independent samples**, use:
 - **Pooled t** when equal variances can be assumed
 - **Welch's t** as the safe default (no equal-variance assumption)
 - **z -test** when both samples are large
- 4 The **F -test** compares two population variances but is sensitive to non-normality
- 5 Always ask: **Are the samples paired or independent?** This is the single most important question for choosing the right test

Looking Ahead

These two-sample tests are the foundation for more advanced methods: ANOVA (comparing > 2 means), regression analysis, and difference-in-differences—all coming in Weeks 12–14.

Every two-sample t -test is equivalent to a **simple linear regression**:

$$Y_i = \beta_0 + \beta_1 D_i + \varepsilon_i$$

where $D_i = 1$ for group 1 and $D_i = 0$ for group 2. Then:

- $\hat{\beta}_0 = \bar{X}_2$ (mean of group 2)
- $\hat{\beta}_1 = \bar{X}_1 - \bar{X}_2$ (difference in means)
- Testing $H_0: \beta_1 = 0$ is **exactly** the two-sample t -test!

This connection is why mastering hypothesis testing now prepares you for regression analysis later.

- Wackerly, Mendenhall & Scheaffer, Ch. 10: Hypothesis Testing—Two Samples
- Hogg, Tanis & Zimmerman, Ch. 6: Tests of Statistical Hypotheses
- Devore & Berk, Ch. 9: Inferences Based on Two Samples
- Wooldridge, *Introductory Econometrics*, Ch. 2: The Simple Regression Model (for the connection between t -tests and regression)